



A133

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :  C12N 15/13, C07K 15/28		A1	(11) International Publication Number: WO 91/10737  (43) International Publication Date: 25 July 1991 (25.07.91)		
(21) International Application Number: PCT/US91/00209  (22) International Filing Date: 10 January 1991 (10.01.91)		(81) Designated States: AT, AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH, CH (European patent), CM (OAPI patent), DE, DE (European patent), DK, DK (European patent), ES, ES (European patent), FI, FR (European patent), GA (OAPI patent), GB, GB (European patent), GR, GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU, LU (European patent), MC, MG, ML (OAPI patent), MR (OAPI patent), MW, NL, NL (European patent), NO, RO, SD, SE, SE (European patent), SN (OAPI patent), SU, TD (OAPI patent), TG (OAPI patent).			
(30) Priority data: 464,350 11 January 1990 (11.01.90) US					
(71) Applicant: MOLECULAR AFFINITIES CORPORATION [US/US]; 291 Broadway, New York, NY 10007 (US).					
(72) Inventor: WIGLER, Michael, H. ; 1 Walden Court, Lloyd Harbor, NY 11743 (US).					
(74) Agents: GRANAHAN, Patricia et al.; Hamilton, Brook, Smith & Reynolds, Two Militia Drive, Lexington, MA 02173 (US).					
(54) Title: PRODUCTION OF ANTIBODIES USING GENE LIBRARIES					
(57) Abstract					
<p>A method of producing libraries of genes encoding antigen-combining molecules or antibodies; a method of producing antigen-combining molecules which does not require an <i>in vivo</i> procedure; a method of obtaining antigen-combining molecules of selected specificity which does not require an <i>in vivo</i> procedure; vectors useful in the present method; and antigen-combining molecules produced by the method. The antigen-combining molecules are useful for the detection, quantitation, purification and neutralization of antigens, as well as for diagnostic, therapeutic and prophylactic purposes.</p>					
<pre> graph TD     A[Construct libraries of genes encoding Ig heavy chain and/or light chain in E. coli vector] --&gt; B[Increase diversity of libraries via random mutagenesis (optional)]     B --&gt; C[Transfect libraries into cultured cells, where they are expressed]     C --&gt; D[Identify cultured cells expressing Ab of desired specificity]     D --&gt; E[Isolate gene(s) encoding Ab of desired specificity and express]   </pre>					

**FOR THE PURPOSES OF INFORMATION ONLY**

**Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.**

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LJ	Liechtenstein	SU	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark				

-1-

Production of antibodies using gene libraries.

Description

Background of the Invention

5        Monoclonal and polyclonal antibodies are useful for a variety of purposes. The precise antigen specificity of antibodies makes them powerful tools that can be used for the detection, quantitation, purification and neutralization of antigens.

10      Polyclonal antibodies are produced in vivo by immunizing animals, such as rabbits and goats, with antigens, bleeding the animals and isolating polyclonal antibody molecules from the blood. Monoclonal antibodies are produced by hybridoma cells, which are made by

15      fusing, in vitro, immortal plasmacytoma cells with antibody producing cells (Kohler, G. and C. Milstein, Nature, 256:495 (1975)) obtained from animals immunized in vivo with antigen.

20      Current methods for producing polyclonal and monoclonal antibodies are limited by several factors. First, methods for producing either polyclonal or monoclonal antibodies require an in vivo immunization step. This can be time consuming and require large amounts of antigen. Second, the repertoire of antibodies expressed

25      in vivo is restricted by physiological processes, such as those which mediate self-tolerance that disable auto-reactive B cells (Goodnow, C.C., et al., Nature, 334:676

-2-

(1988); Goodnow, J.W., Basic and Clinical Immunology, Ed. 5, Los Altos, CA, Large Medical Publications (1984); Young, C.R., Molecular Immunology, New York, Marcel Dekker (1984)). Third, although antibodies can exist in millions of different forms, each with its own unique binding site for antigen, antibody diversity is restricted by genetic mechanisms for generating antibody diversity (Honjo, T., Ann. Rev. Immunol., 1:499 (1983); Tonegawa, S., Nature:302:575 (1983)). Fourth, not all the antibody molecules which can be generated will be generated in a given animal. As a result, raising high affinity antibodies to a given antigen can be very time consuming and can often fail. Fifth, the production of human antibodies of desired specificity is very problematical.

A method of producing antibodies which avoids the limitations of presently-available methods, such as the requirement for immunization of an animal and in vivo steps, would be very useful, particularly if it made it possible to produce a wider range of antibody types than can be made using presently-available techniques and if it made it possible to produce human antibody types.

Disclosure of the Invention

The present invention relates to a method of producing libraries of genes encoding antigen-combining molecules or antibodies; a method of producing antigen-combining molecules, also referred to as antibodies, which does not require an in vivo procedure, as is required by presently-available methods; a method of obtaining antigen-combining molecules (antibodies) of

- 3 -

selected or defined specificity which does not require an in vivo procedure; vectors useful in the present method and antibodies produced or obtained by the method.

The present invention relates to an in vitro process  
05 for synthesizing DNA encoding families of antigen-combining molecules or proteins. In this process, DNA containing genes encoding antigen-combining molecules is obtained and combined with oligonucleotides which are homologous to regions of the genes which are conserved.  
10 Sequence-specific gene amplification is then carried out using the DNA containing genes encoding antigen-combining proteins as template and the homologous oligonucleotides as primers.

This invention also relates to a method of creating  
15 diverse libraries of DNAs encoding families of antigen-combining proteins by cloning the product of the in vitro process for synthesizing DNA, described in the preceding paragraph, into an appropriate vector (e.g., a plasmid, viral or retroviral vector).

20 The subject invention provides an alternative method for the production of antigen-combining molecules, which are useful affinity reagents for the detection and neutralization of antigens and the delivery of molecules to antigenic sites. The claimed method differs from  
25 production of polyclonal antibody molecules derived by immunization of live animals and from production of monoclonal antibody molecules through the use of hybridoma cell lines in that it does not require an in vivo immunization step, as do presently available methods.  
30 Rather, diverse libraries of genes which encode antigen-combining sites comprising a significant proportion of an

-4-

animal's repertoire of antibody combining sites are made, as described in detail herein. These genes are expressed in living cells, from which molecules of desired antigenic selectivity can be isolated and purified for  
05 various uses.

Antigen-combining molecules are produced by the present method in the following manner, which is described in greater detail below. Initially, a library of antibody genes which includes a set of variable 10 regions encoding a large, diverse and random group of specificities derived from animal or human immunoglobulins is produced by amplifying or cloning diverse genomic fragments or cDNAs of antibody mRNAs found in antibody-producing tissue.

15 In an optional step, the diversity of the resulting libraries can be increased by means of random mutagenesis. The gene libraries are introduced into cultured host cells, which may be eukaryotic or prokaryotic, in which they are expressed. Genes encoding antibodies of 20 desired antigenic specificity are identified, using a method described herein or known techniques, isolated and expressed in quantities in appropriate host cells, from which the encoded antibody can be purified.

25 Specifically, a library of genes encoding immunoglobulin heavy chain regions and a library of genes encoding immunoglobulin light chain regions are constructed. This is carried out by obtaining antibody-encoding DNA, which is either genomic fragments or cDNAs of antibody mRNAs, amplifying or cloning the fragments or 30 cDNAs; and introducing them into a standard framework antibody gene vector, which is used to introduce the

- 5 -

antibody-encoding DNA into cells in which the DNA is expressed. The vector includes a framework gene encoding a protein, such as a gene encoding an antibody heavy chain or an antibody light chain which can be of any origin (human, non-human) and can be derived from any of a number of existing DNAs encoding heavy chain immunoglobulins or light chain immunoglobulins. Such vectors are also a subject of the present invention and are described in greater detail in a subsequent section.

05 Genes from one or both of the libraries are introduced into appropriate host cells, in which the genes are expressed, resulting in production of a wide variety of antigen-combining molecules.

10

Genes encoding antigen-combining molecules of desired specificity are identified by identifying cells producing antigen-combining molecules which react with a selected antigen and then obtaining the genes of interest. The genes of interest can subsequently be introduced into an appropriate host cell (or can be

15 further modified and then introduced into an appropriate host cell) for further production of antigen-combining molecules, which can be purified and used for the same purposes, for which conventionally-produced antibodies are used.

20

25 Through use of the method described, it is possible to produce antigen-combining molecules which are of wider diversity than are antibodies available as a result of known methods; novel antigen-combining molecules with a diverse range of specificities and affinities and

30 antigen-combining molecules which are predominantly human in origin. Such antigen-combining molecules are a

- 6 -

subject of the present invention and can be used clinically for diagnostic, therapeutic and prophylactic purposes, as well as in research contexts, and for other purposes.

05 Brief Description of the Drawings

Figure 1 is a schematic representation of the method of the present invention by which antigen-combining molecules, or antibodies, are produced.

10 Figure 2 is a schematic representation of amplification or cloning of IgM heavy chain variable region DNA from mRNA, using the polymerase chain reaction.

Panel A shows the relevant regions of the poly adenylated mRNA encoding the secreted form of the IgM heavy chain. S denotes the sequences encoding the signal peptide which causes the nascent peptide to cross the plasma membrane. V, D and J together comprise the variable region.  $C_H^1$ ,  $C_H^2$ , and  $C_H^3$  are the three constant domains of  $C\mu$ . Hinge encodes the hinge region. C, B and Z are oligonucleotide PCR primers (discussed below).

20 Panel B shows the reverse transcript DNA product of the mRNA primed by oligonucleotide Z, with the addition of poly-dC by terminal transferase at the 3' end.

Panel C is a schematic representation of the annealing of primer A to the reverse transcript DNA.

25 Panel D shows the final double stranded DNA PCR product made utilizing primers A and B.

Panel E shows the product of PCR annealed to primer C.

Panel F is a blowup of Panel E, showing in greater detail the structure of primer C. Primer C consists of two

- 7 -

parts: a 3' part complementary to IgM heavy chain mRNA as shown, and a 5' part which contains restriction site RE2 and spacer.

05 Panel G shows the final double stranded DNA PCR product made utilizing primers A and C and the product of the previous PCR (depicted in D) as template. The S, V, D, J regions are again depicted.

10 Figure 3 is a schematic representation of the heavy chain framework vector pFHC. The circular plasmid (above) is depicted linearized (below) and its relevant components are shown: animal cell antibiotic resistance marker; bacterial replication origin; bacterial cell antibiotic resistance marker;  $C\mu$  enhancer; LTR containing the viral promoter from the Moloney MLV retrovirus DNA; 15 PCR primer (D); cDNA cloning site containing restriction endonuclease sites, RE1 and RE2, separated by spacer DNA;  $C\mu$  exons; and poly A addition and termination sequences derived from the  $C\mu$  gene or having the same sequence as the  $C\mu$  gene.

20 Figure 4 depicts a nucleotide sequence of the  $C_H^1$  exon of the  $C\mu$  gene, and its encoded amino acid sequence (Panel A). The nucleotide coordinate numbers are listed above the line of nucleotide sequences. Panel B depicts the N-doped sequence, as defined in the text.

25 Detailed Description of the Invention

The present invention provides a method of producing antigen-combining molecules (or antibodies) which does not require an in vivo immunization procedure and which makes it possible to produce antigen-combining molecules

- 8 -

with far greater diversity than is shown by antibodies produced by currently-available techniques.

The present invention relates to a method of producing libraries of genes encoding antigen-combining molecules (antibody proteins) with diverse antigen-combining specificities; a method of producing such antigen-combining molecules, antigen-combining molecules produced by the method and vectors useful in the method. The following is a description of generation 05 of such libraries, of the present method of producing antigen-combining molecules of selected specificity and 10 of vectors useful in producing antigen-combining molecules of the present invention.

As described below, the process makes use of 15 techniques which are known to those of skill in the art and can be applied as described herein to produce and identify antigen-combining molecules of desired antigenic specificity: the polymerase chain reaction (PCR), to amplify and clone diverse cDNAs encoding antibody mRNAs 20 found in antibody-producing tissue; mutagenesis protocols to further increase the diversity of these cDNAs; gene transfer protocols to introduce antibody genes into cultured (prokaryotic and eukaryotic) cells for the purpose of expressing them; and screening protocols to 25 detect genes encoding antibodies of the desired antigenic specificity. A general outline of the present method is represented in Figure 1.

- 9 -

Construction of Library of Genes Encoding  
Antigen-Combining Molecules

05 A key step in the production of antigen-combining molecules by the present method is the construction of a "library" of antibody genes which include "variable" regions encoding a large, diverse, but random set of specificities. The library can be of human or non-human origin and is constructed as follows:

10 Initially, genomic DNA encoding antibodies or cDNAs of antibody mRNA (referred to as antibody-encoding DNA) is obtained. This DNA can be obtained from any source of antibody-producing cells, such as spleen cells, peripheral blood cells, lymph nodes, inflammatory tissue cells and bone marrow cells. It can also be obtained 15 from a genomic library or cDNA library of B cells. The antibody-producing cells can be of human or non-human origin; genomic DNA or mRNA can be obtained directly from the tissue (i.e., without previous treatment to remove cells which do not produce antibody) or can be obtained 20 after the tissue has been treated to increase concentration of antibody-producing cells or to select a particular type(s) of antibody-producing cells (i.e., treated to enrich the content of antibody-producing cells). Antibody-producing cells can be stimulated by an 25 agent which stimulates antibody mRNA production (e.g., lipopolysaccharide) before DNA is obtained.

30 Antibody-encoding DNA is amplified and cloned using a known technique, such as the PCR using appropriately-selected primers, in order to produce sufficient quantities of the DNA and to modify the DNA in such a manner

-10-

(e.g., by addition of appropriate restriction sites) that it can be introduced as an insert into an E.coli cloning vector. This cloning vector can serve as the expression vector or the inserts can later be introduced into an 05 expression vector, such as the framework antibody gene vector described below. Amplified and cloned DNA can be further diversified, using mutagenesis, such as PCR, in order to produce a greater diversity or wider repertoire 10 of antigen-binding molecules, as well as novel antigen-binding molecules.

Cloned antibody-encoding DNA is introduced into an expression vector, such as the framework antibody gene vector of the present invention, which can be a plasmid, viral or retroviral vector. Cloned antibody-encoding DNA 15 is inserted into the vector in such a manner that the cloned DNA will be expressed as protein in appropriate host cells. It is essential that the expression vector used make it possible for the DNA insert to be expressed as a protein in the host cell. One expression vector 20 useful in the present method is referred to as the framework antibody gene vector. Vectors useful in the present method contain antibody constant region or portions thereof in such a manner that when amplified DNA 25 is inserted, the vector expresses a chimeric gene product comprising a variable region and a constant region in proper register. The two regions present in the chimeric gene product can be from the same type of immunoglobulin molecule or from two different types of immunoglobulin molecules.

30 These libraries of antibody-encoding genes are then expressed in cultured cells, which can be eukaryotic or

-11-

prokaryotic. The libraries can be introduced into host cells separately or together. Introduction of the antibody-encoding DNA in vitro into host cells (by infection, transformation or transfection) is carried out 05 using known techniques, such as electroporation, protoplast fusion or calcium phosphate co-precipitation. If only one library is introduced into a host cell, the host cell will generally be one which makes the other antibody chain, thus making it possible to produce 10 complete/functional antigen-binding molecules. For example, if a heavy chain library produced by the present method is introduced into host cells, the host cells will generally be cultured cells, such as myeloma cells or E. coli, which naturally produce the other (i.e., light) 15 chain of the immunoglobulin or are engineered to do so. Alternatively, both libraries can be introduced into appropriate host cells, either simultaneously or sequentially.

Host cells in which the antibody-encoding DNA is 20 expressed can be eukaryotic or prokaryotic. They can be immortalized cultured animal cells, such as a myeloma cell line which has been shown to efficiently express and secrete introduced immunoglobulin genes (Morrison, S.L., et al., Ann. N.Y. Acad. Sci., 507:187 (1987); Kohler, G. 25 and C. Milstein, Eur. J. Immunol., 6:511 (1976); Oi, V.T., et al., Immunoglobulin Gene Expression in Transformed Lymphoid Cells, 80:825 (1983); Davis, A.C. and M.J. Shulman, Immunol. Today, 10:119 (1989)). One 30 host cell which can be used to express the antibody-encoding DNA is the J558L cell line or the SP2/0 cell line.

-12-

Cells expressing antigen-combining molecules with a desired specificity for a given antigen can then be selected by a variety of means, such as testing for reactivity with a selected antigen using nitrocellulose layering. The antibodies identified thereby can be of human origin, nonhuman origin or a combination of both. That is, all or some of the components (e.g., heavy chain, light chain, variable regions, constant regions) can be encoded by DNA of human or nonhuman origin, which, when expressed produces the encoded chimeric protein which, in turn, may be human, nonhuman or a combination of both. In such antigen-combining molecules, all or some of the regions (e.g., heavy and light chain variable and constant regions) are referred to as being of human origin or of nonhuman origin, based on the source of the DNA encoding the antigen-combining molecule region in question. For example, in the case in which DNA encoding mouse heavy chain variable region is expressed in host cells, the resulting antigen-combining molecule has a heavy chain variable region of mouse origin. Antibodies produced may be used for such purposes as drug delivery, tumor imaging and other therapeutic, diagnostic and prophylactic uses.

Once antibodies of a desired binding specificity are obtained, their genes may be isolated and further mutagenized to create additional antigen combining diversity or antibodies of higher affinity for antigen.

-13-

Construction of Immunoglobulin Heavy Chain Gene Library  
and Production of Encoded Antigen-Binding Molecules

The following is a detailed description of a specific experimental protocol which embodies the concepts described above. Although the following is a description of one particular embodiment, the same procedures can be used to produce libraries in which the immunoglobulin and the heavy chain class are different or in which light chain genes are amplified and cloned. The present invention is not intended to be limited to this example. In the embodiment presented below, a diverse heavy chain gene library is constructed. Using the principles described in relation to the heavy chain gene library, a diverse light chain gene library is also constructed. These are co-expressed in an immortal tumor cell capable of producing antibodies, such as plasma-cytoma cells or myeloma cells. Cells expressing antibody reactive to antigen are identified by a nitrocellulose filter overlay and antibody is prepared from cells identified as expressing it. As described in a subsequent section, there are alternative methods of library construction, other expression systems which can be used, and alternative selection systems for identifying antibody-producing cells or viruses.

Step 1 in this specific protocol is construction of libraries of genes in E. coli which encode immunoglobulin heavy chains. This is followed by the use of random mutagenesis to increase the diversity of the library, which is an optional procedure. Step 2 is introduction of the library, by transfection, into myeloma cells.

-14-

Step 3 is identification of myeloma cells expressing antibody with the desired specificity, using the nitrocellulose filter overlay technique or techniques known to those of skill in the art. Step 4 is isolation 05 of the gene(s) encoding the antibody with the desired specificity and their expression in appropriate host cells, to produce antigen-combining fragments useful for a variety of purposes.

Construction

10 One key step in construction of the library of cDNAs encoding the variable region of mouse heavy chain genes is construction of an E. coli plasmid vector, designated pFHC. pFHC contains a "framework" gene, which can be any antibody heavy chain and serves as a site into which 15 the amplified cloned gene product (genomic DNA or cDNA of antibody mRNAs) is introduced. pFHC is useful as a vector for this purpose because it contains RE1 and RE2 cloning sites. Other vectors which include a framework gene and other cloning sites can be used for this purpose 20 as well. The framework gene includes a transcriptional promoter (e.g., a powerful promoter, such as a Moloney LTR (Mulligan, R.C., In Experimental Manipulation of Gene Expression, New York Adademic Press, p. 155 (1983)) and a C $\mu$  chain transcriptional enhancer to increase the level 25 of transcriptions from the promoter (Gillies, S.D., et al., Cell, 33:717 (1983), a cloning site containing RE1 and RE2; part of the C $\mu$  heavy chain gene encoding secreted protein; and poly A addition and termination sequences (Figure 3). The framework antibody gene vector 30 of the present invention (pFHC) also includes a

-15-

selectable marker (e.g., an antibiotic resistance gene such as the neomycin resistance gene, neo<sup>R</sup>) for animal cells; sequences for bacterial replication (ori); and a selectable marker (e.g., the ampicillin resistance gene, Amp<sup>R</sup>) for bacterial cells. The framework gene can be of any origin (human, non-human), and can derive from any one of a number of existing DNAs encoding heavy chain immunoglobulins (Tucker, P.W., et al., Science, 206:1299 (1979); Honjo, T., et al., Cell, 18:559 (1979); Bothwell, 05 A.L.M., et al., Cell, 24:625 (1981); Liu, A.Y., et al., Gene, 54:33 (1987); Kawakami, T., et al., Nuc. Acids. Res., 8:3933 (1980)). In this embodiment, the vector retains the introns between the C<sub>H</sub>1, hinge, C<sub>H</sub>2 and C<sub>H</sub>3 10 exons. The "variable region" of the gene, which includes the V, D and J regions of the antibody heavy chain and which encodes the antigen binding site, is deleted and replaced with two consecutive restriction endonuclease 15 cloning sites, RE1 and RE2. The restriction endonuclease site RE1 occurs just 3' to the LTR promoter and the restriction endonuclease site RE2 occurs within the 20 constant region just 3' to the J region (see Figure 3).

Another key step in the production of antigen-combining molecules in this embodiment of the present invention is construction in an E. coli vector of a 25 library of cDNAs encoding the variable region of mouse immunoglobulin genes. In this embodiment, the pFHC vector, which includes cloning sites designated RE1 and RE2, is used for cloning heavy chain variable regions, although any cloning vector with cloning sites having the 30 same or similar characteristics (described below) can be used. Similarly, a light chain vector can be designed,

-16-

using the above described procedures and procedures known to a person of ordinary skill in the art.

05 In this embodiment, non-immune mouse spleens are used as the starting material. mRNA is prepared directly from the spleen or from spleen processed in such a manner that it is enriched for resting B cells. Enrichment of tissue results in a more uniform representation of antibody diversity in the starting materials.

10 Lymphocytes can be purified from spleen using ficoll gradients (Boyum, A., Scand. J. of Clinical Invest., 21:77 (1968)). B cells are separated from other cells (e.g., T cells) by panning with anti-IgM coated dishes (Wysocki, L.J. and V.L. Sato, Proc. Natl. Acad. Sci., 75:2844 (1978)). Because activated cells express the 15 IL-2 receptor but resting B cells do not, resting B cells can be separated yet further from activated cells by panning. Further purification by size fractionation on a Cell Sorter results in a fairly homogeneous population of resting B cells.

20 Poly A+ mRNA from total mouse spleen is prepared according to published methods (Sambrook, J., et al., Molecular Cloning: A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989)). Production of antibody mRNA can first be 25 stimulated by lipopolysaccharide (LPS) (Andersson, J.A., et al., J. Exp. Med., 145:1511 (1977)). First strand cDNA is prepared to this mRNA population using as primer an oligonucleotide, Z, which is complementary to C $\mu$  in the C $\mu$ 1 region 3' to J. This primer is designated Z in 30 Figure 2. First strand cDNA is then elongated by the terminal transferase reaction with dCTP to form a poly dC

-17-

tail (Sambrook, J., *et al.*, Molecular Cloning: A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989)).

This DNA product is then used as template in a  
05 polymerase chain reaction (PCR) to amplify cDNAs encoding  
antibody variable regions (Saiki, R.K., *et al.*, Science,  
239:487 (1988); Ohara, O., *et al.*, Proc. Natl. Acad. Sci. USA, 86:5673 (1989)). Initially, PCR is carried out with  
two primers: primer A and primer B, as represented in  
10 Figure 2. Primer A contains the RE1 site at its 5' end,  
followed by poly dG. Primer B is complementary to the  
constant ( $C_H^1$ ) region of the  $C_{\mu}$  gene, 3' to the J region  
and 5' to primer Z (see Figure 2). Primer B is  
complementary to all  $C_{\mu}$  genes, which encode the heavy  
15 chain of molecules of the IgM class, the Ig class  
expressed by all B cell clones prior to class switching  
(Schimizu, A. and T. Honjo, Cell, 36:801-803 (1984)) and  
present in resting B cells. The resultant PCR product  
includes a significant proportion of cDNAs encompassing  
20 the various  $V_H$  regions expressed as IgM in the mouse.  
(The use of other primers complementary to the cDNA genes  
encoding the constant regions of other immunoglobulin  
heavy chains can be used in parallel reactions to obtain  
the variable regions expressed on these molecules, but  
25 for simplicity these are not described).

Next, the product of the first PCR procedure is used  
again for PCR with primer A and primer C. Primer C, like  
primer B, is complementary to the  $C_{\mu}$  gene 3' to J and  
just 5' to primer B (see Figure 2). Primer C contains  
30 the RE2 site at its 5' end. The RE2 sequence is chosen  
in such a manner that when it is incorporated into the

-18-

framework vector, no alteration of coding sequence of the C $\mu$  chain occurs (See Figures 2 and 3). This method of amplifying C $\mu$  cDNAs, referred to as unidirectional nested PCR, incorporates the idea of nested primers for cloning 05 a gene when the nucleotide sequence of only one region of the gene is known (Ohara, O., et al., Proc. Natl. Acad. Sci. USA, 86:5673 (1989)). The PCR product is then cleaved with restriction enzymes RE1 and RE2 and cloned into the RE1 and RE2 sites of the pFHC vector (described 10 below). The sequence of primers and of RE1 and RE2 sites are selected so that when the PCR product is cloned into these sites, the sites are recreated and the cloned antibody gene fragments are brought back into the proper frame with respect to the framework immunoglobulin gene 15 present in pFHC. This results in creation of a C $\mu$  minigene which lacks the intron normally present between J and the C<sub>H</sub>1 region of C $\mu$  (See Figure 3). These procedures result in production of the heavy chain library used to produce antigen-binding molecules of the 20 present invention, as described further below.

Optionally, diversity of the heavy chain variable region is increased by random mutagenesis, using techniques known to those of skill in the art.

For example, the library produced as described above 25 is amplified again, using PCR under conditions of limiting nucleotide concentration. Such conditions are known to increase the infidelity of the polymerization and result in production of mutant products. Primers useful for this reaction are Primers C and D, as 30 represented in Figures 2 and 3. Primer D derives from pFHC just 5' to RE1. The PCR product, after cleavage

-19-

with RE1 and RE2, is recloned into the framework vector pFHC. To the extent that mutation affects codons of the antigen binding region, this procedure increases the diversity of the binding domains. For example, if the 05 starter library has a complexity of  $10^6$  elements, and an average of one mutation is introduced per complementarity determining region, and it is assumed that the complementarity determining region is 40 amino acids in size and that any of six amino acid substitutions can 10 occur at a mutated codon, the diversity of the library can be increased by a factor of about  $40 \times 6$ , or 240, for single amino acid changes and  $240 \times 240$ , or about  $6 \times 10^4$ , for double amino acid changes, yielding a final 15 diversity of approximately  $10^{11}$ . This is considered to be in the range of the diversity of antibodies which animals produce (Tonegawa, S., Nature, 302:575 (1983)). Even greater diversity can be generated by the random combination of H and L chains, the result of co-expression 20 in host cells (see below). It is, thus, theoretically possible to generate a more diverse antibody library in vitro than can be generated in vivo. This library of genes is called the "high diversity" heavy chain library. It may be propagated indefinitely in E. coli. A high diversity light chain library can be 25 prepared similarly.

The framework vector for the light chain library, designated pFLC, includes components similar to those in the vector for the heavy chain library: the enhancer, promoter, a bacterial selectable marker, an animal 30 selectable marker, bacterial origin of replication and light chain exons encoding the constant regions. For

-20-

pFLC, the animal selectable marker should differ from the animal selectable in pFHC. For example, if pFHC contains neo<sup>R</sup>, pFLC can contain Eco gpt.

A light chain library, which contains diverse light  
05 chain fragments, is prepared as described above for construction of the heavy chain library. In constructing the light chain library, the primers used are different from those described above for heavy chain library construction. In this instance, the primers are  
10 complementary to light chain mRNA encoding constant regions. The framework vector contains the light chain constant region exons.

Introduction of the Library of Immunoglobulin Chain Genes into Immortalized Animal Cells

15 The library of immunoglobulin chain genes produced as described is subsequently introduced into a line of immortalized cultured animal cells, referred to as the "host" cells, in which the genes in the library are expressed. Particularly useful for this purpose are  
20 plasmacytoma cell lines or myeloma cell lines which have been shown to efficiently express and secrete introduced immunoglobulin genes (Morrison, S.L., et al., Ann. N.Y. Acad. Sci., 507:187 (1987); Kohler, G. and C. Milstein, Eur. J. Immunol., 6:511 (1976); Galfre and C. Milstein,  
25 Methods Enzymol., 73:3 (1981); Davis, A.C. and M.J. Shulman, Immunol. Today, 10:119 (1989)). For example, the J558L cell line can be cotransfected using electroporation or protoplast fusion (Morrison, S.L., et al., Ann. N.Y. Acad. Sci., 507:187 (1987)) and transfected

- 21 -

cells selected on the basis of auxotrophic markers present on light and heavy chain libraries.

As a result of cotransformation and selection for markers on both light chain and heavy chain vectors, most 05 transformed host cells will express several copies of immunoglobulin heavy and light chains from the diverse library, and will express chimeric antibodies (antibodies encoded by all or part of two or more genes) (Nisonoff, A., *et al.*, In The Antibody Molecule, Academic Press, NY 10 p. 238 (1975)). These chimeric antibodies are of two types: those in which one chain is encoded by a host cell gene and the other chain is encoded by an exogenously introduced antibody gene and those in which both the light and the heavy chain are encoded by an exogenous 15 antibody gene. Both types of antibodies will be secreted. A library of cells producing antibodies of diverse specificities is produced as a result. The library of cells can be stored and maintained indefinitely by continuous culture and/or by freezing. A 20 virtually unlimited number of cells can be obtained by this process.

Isolation of Cells Producing Antigen-Binding Molecules of Selected Specificity

25 Cells producing antigen-binding molecules of selected specificity (i.e., which bind to a selected antigen) can be identified and isolated using nitrocellulose filter layering or known techniques. The same methods employed to identify and isolate hybridoma 30 cells producing a desired antibody can be used: cells are pooled and the supernatants tested for reactivity

- 22 -

with antigen (Harlow, E. and D. Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, N.Y., p. 283 (1988). Subsequently, individual clones of cells are identified, using known techniques. A preferred 05 method for identification and isolation of cells makes use of nitrocellulose filter overlays, which allow the screening of a large number of cells. Cells from the library of transfected myeloma cells are seeded in 10 cm<sup>2</sup> petri dishes in soft agar (Cook, W.D. and M.D. Scharff, 10 PNAS, 74:5687 (1977); Paige, C.J., et al., Methods in Enzymol., 150:257 (1987)) at a density of 10<sup>4</sup> colony forming units, and allowed to form small colonies (approximately 300 cells). A large number of dishes (>100) may be so seeded. Cells are then overlayed with a 15 thin film of agarose (<1mm) and the agarose is allowed to harden. The agarose contains culture medium without serum. Nitrocellulose filters (or other protein-binding filters) are layered on top of the agarose, and the dishes are incubated overnight. During this time, 20 antibodies secreted by the cells will diffuse through the agarose and adhere to the nitrocellulose filters. The nitrocellulose filters are keyed to the underlying plate and removed for processing.

The method for processing nitrocellulose filters is 25 identical to the methods used for Western blotting (Harlow, E. and D. Lane, Antibodies: Laboratory Manual, Cold Spring Harbor, N.Y., p. 283 (1988)). The antibody molecules are adsorbed to the nitrocellulose filter. The filters, as prepared above, are then blocked. The 30 desired antigen, for example, keyhole lymphet hemocyanin (KLH), which has been iodinated with radioactive <sup>125</sup>I, is

- 23 -

then applied in Western blotting buffers to the filters. (Other, non radiographic methods can be used for detection). After incubation, the filters are washed and dried and used to expose autoradiography film according to standard procedures. Where the filters have adsorbed antibody molecules which are capable of binding KLH, the autoradiography film will be exposed. Cells expressing the KLH reactive antibody can be identified by determining the location on the dish corresponding to an exposed filter; cells identified in this manner can be isolated using known techniques. Cells which are isolated from a region of the dish can then be rescreened, to insure the isolation of the clone of antigen-binding molecule-producing cells.

15 Isolation of Genes Encoding Antigen-Binding Molecules of Selected Specificity and Purification of Encoded Antigen-Binding Molecules

The gene(s) encoding an antigen-binding molecule of selected specificity can be isolated. This can be carried out, for example, as follows: primers D and C (see Figures 2 and 3) are used in a polymerase chain reaction, to produce all the heavy chain variable region genes introduced into the candidate host cell from the library. These genes are cloned again in the framework vector pFHC at the RE1 and RE2 sites. Similarly, all the light chain regions introduced into the host cell from the library are cloned into the light chain vector, pFLC. Members of the family of vectors so obtained are then transformed pairwise into myeloma cells, which are tested for the ability to produce and secrete the antibody with

-24-

the desired selectivity. Purification of the antibody from these cells can then be accomplished using standard procedures (Johnstone, A. and R. Thorpe, Immunochem. in Practice, Blackwell Scientific, Oxford, p. 27 (1982);

05 Harlow, E. and D. Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, N.Y., p. 283 (1988)).

Alteration of Affinity of Antigen-Binding Molecules

It is also possible to produce antigen-binding molecules whose affinity for a selected antigen is altered (e.g., different from the affinity of a corresponding antigen-binding molecule produced by the present method). This can be carried out, for example, to increase the affinity of an antigen-binding molecule by randomly mutagenizing the genes isolated as described above using previously-described mutagenesis methods.

Alternatively, the variable region of antigen-binding molecule-encoding genes can be sequenced and site directed mutagenesis performed to mutate the complementarity determining regions (CDR) (Kabat, E.A., J. Immunol., 141:S 25-36 (1988)). Both processes result in production of a sublibrary of genes which can be screened for antigen-binding molecules of higher affinity or of altered affinity after the genes are expressed in myeloma cells.

25 Alternative Materials and Procedures for Use in the Present Method

In addition to those described above for use in the method of the present invention, other materials (e.g., starting materials, primers) and procedures can be used

- 25 -

in carrying out the method. For example, use of PCR technology to clone a large collection of cDNA genes encoding variable regions of heavy chains has been described above. Although primers from the  $C\mu$  class were 05 described as being used in unidirectional nested PCR, the present invention is not limited to these conditions. For example, primers from any of the other heavy chain classes ( $C\gamma_3$ ,  $C\gamma_1$ ,  $C\gamma_2b$ ,  $C\alpha$  for example) or from light chains can be used.  $C\mu$  was described as of particular 10 use because of the fact that the entire repertoire of heavy chain variable regions are initially expressed as IgM. Only following heavy-chain class switching are these variable regions expressed with a heavy chain of a different class (Shimizu, A. and T. Honjo, Cell, 15 36:801-803 (1984)). In addition, the predominant population of B cells in nonimmune spleen cells is IgM<sup>+</sup>-cells (Cooper, M.D. and P. Burrows, In Immunoglobulin Genes, Academic Press, N.Y. p. 1 (1989)). Although unidirectional nested PCR amplification is 20 described above, other PCR procedures, as well as other DNA amplification techniques can be used to amplify DNA as needed in the present invention. For example, bidirectional PCR amplification of antibody variable regions can be carried out. This approach requires use 25 of multiple degenerate 5' primers (Orlandi, R., et al., Proc. Natl. Acad. Sci. USA, 86:3833 (1989); Sastry, L., et al., Proc. Natl. Acad. Sci. USA, 86:5728 (1989)). Bidirectional amplification may not pick up the same full 30 diversity of genes as can be expected from unidirectional PCR.

-26-

In addition, methods of introducing further diversity into the antibody library other than the method for random mutagenesis utilizing PCR described above can be used. Other methods of random mutagenesis, such as 05 that described by Sambrook, *et al.* (Sambrook, J., *et al.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)) can be used, as can direct mutagenesis of the complementarity determining regions (CDRs).

10 Framework vectors other than one using a mouse C $\mu$  heavy chain constant region, which contains the C $\mu$  enhancer and introns and a viral promoter (described previously) can be used for inserting the products of PCR. The vectors described were chosen for their 15 subsequent use in the expression of the antibody genes, but any eukaryotic or prokaryotic cloning vector could be used to create a library of diverse cDNA genes encoding variable regions of antibody molecules. The inserts from this vector could be transferred to any number of 20 expression vectors. For example, other framework vectors which include intronless genes can be constructed, as can other heavy chain constant regions. In addition to plasmid vectors, viral vectors or retroviral vectors can be used to introduce genes into myeloma cells.

25 The source for antibody molecule mRNAs can also be varied. Purified resting B lymphocytes from mouse nonimmunized spleen are described above as such a source. However, total spleens (immunized or not) from other animals, including humans, can be used, as can any source 30 of antibody-producing cells (e.g., peripheral blood, lymph nodes, inflammatory tissue, bone marrow).

-27-

Introduction of H and L chain gene DNA into myeloma cells using cotransformation by electroporation or protoplast fusion methods is described above (Morrison, S.L. and V.T. Oi, Adv. Immunol., 44:65 (1989)). However, 05 any means by which DNA can be introduced into living cells in vivo can be used, provided that it does not significantly interfere with the ability of the transformed cells to express the introduced DNA. In fact, a method other than cotransformation, can be used. 10 Cotransfection was chosen for its simplicity, and because both the H and L chains can be introduced into myeloma cells. It may be possible to introduce only the H chain into myeloma cells. Moreover, the H chain itself in many cases carries sufficient binding affinity for antigen. 15 However, other methods can also be used. For example, retroviral infection may be used. Replication-incompetent retroviral vectors can be readily constructed which can be packaged into infective particles by helper cells (Mann, R., et al., Cell, 33:153-159 (1903)). Viral 20 titers of  $10^5$  infectious units per ml. can be achieved, making possible the transfer of very large numbers of genes, into myeloma cells.

Further increases in the diversity of antibody-producing cells than results from the method described 25 above can be generated if light and heavy chain genes are introduced separately into myeloma cells. Light chain genes can be introduced into one set of myeloma cells with one selectable marker, and heavy chains into another set of cells with a different selectable marker. Myeloma 30 cells containing and expressing both H and L chains could then be generated by the highly efficient process of

- 28 -

polyethylene glycol mediated cell fusion (Pontecorvo, G., Somatic Cell Genetics, 1:397 (1975)). Thus, a method of screening diverse libraries of antibody genes using animal cells is not limited by the number of cells which 05 can be generated, but by the number of cells which can be screened.

Methods of identifying antigen-binding molecule-expressing cells expressing an antigen-binding molecule of selected specificity other than the nitrocellulose 10 filter overlay technique described above can be used. An important characteristic of any method is that it be useful to screen large numbers of different antibodies. With the nitrocellulose filter overlay technique, for example, if 300 dishes are prepared and  $10^4$  independent 15 transformed host cells per dish are screened, and if, on average, each cell produces ten different antibody molecules, then  $300 \times 10^4 \times 3$ , or about  $10^7$  different antibodies can be screened at once. However, if the antibody molecules can be displayed on the cell surface, 20 still larger numbers of cells can be screened using affinity matrices to pre-enrich for antigen-binding cells. There are immortal B cell lines, such as BCL<sub>1</sub>B<sub>1</sub>, which will express IgM both on the cell surface and as a secreted form (Granowicz, E.S., et al., J. Immunol., 25 125:976 (1980)). If such cells are infected by retroviral vectors containing the terminal C $\mu$  exons, the infected cells will likely produce both secreted and membrane bound forms of IgM (Webb, C.F., et al., J. Immunol., 143:3934-3939 (1989)). Still other methods can 30 be used to detect antibody production. If the host cell is E. coli, a nitrocellulose overlay is possible, and

-29-

such methods have been frequently used to detect E. coli producing particular proteins (Sambrook, J., et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

05 (1989)). Other methods of detection are possible and one in particular, which involves the concept of "viral coating", is discussed below.

Viral coating can be used as a means of identifying viruses encoding antigen-combining molecules. In this 10 method, a viral vector is used to direct the synthesis of diverse antibody molecules. Upon lytic infection of host cells, and subsequent cell lysis, the virus becomes "coated" with the antibody product it directs. That is, the antibody molecule becomes physically linked to the 15 outside of a mature virus particle, which can direct its synthesis. Methods for viral coating are described below. Viruses coated by antibody can be physically selected on the basis of their affinity to antigen which is attached to a solid support. The number of particles 20 which can be screened using this approach is well in excess of  $10^9$  and it is possible that  $10^{11}$  different antibody genes could be screened in this manner. In one embodiment, an affinity matrix containing antigen used to purify those viruses encoding antibody molecules with 25 affinity to antigen and which coat the surface of the virus which encodes those antibodies is used.

One method of viral coating is as follows: A diverse library of bacteriophage  $\lambda$  encoding parts of antibody molecules that are expressed in infected E. coli 30 and which retain the ability to bind antigens is created,

-30-

using known techniques (Orlandi, R., et al., Proc. Natl. Acad. Sci. USA, 86:3833 (1989); Huse, W.D., et al., Science, 246:1275 (1989); Better, M., et al., Science, 240:1041 (1988); Skerra, A. and A. Pluckthorn, Science, 240:1038 (1988)). Bacteria infected with phage are embedded in a thin film of semisolid agar. Greater than  $10^7$  infected bacteria may be plated in the presence of an excess of uninfected bacteria in a volume of 1 ml of agar and spread over a  $10\text{ cm}^2$  surface. The agar contains 10 monovalent antibody "A" (Parham, P., In Handbook of Experimental Immunology: Immunochem., Blackwell Scientific Publishers, Cambridge, MA, pp. 14.1-14.23 (1986)), which can bind the  $\lambda$  coat proteins and which has been chemically coupled to monovalent antibody "B", which 15 can bind an epitope on all viral directed antibody molecules. Monovalent antibodies are used to prevent the crosslinking of viral particles. Upon lytic burst, progeny phage particles become effectively cross linked to the antibody molecule they encode. Because lysis 20 occurs in semisolid medium, in which diffusion is slow, cross linking between a given phage and the antibody encoded by another phage is minimized. A nitrocellulose filter (or other protein binding filter) is prepared as an affinity matrix by adsorbing the desired antigen. The 25 filter is then blocked so that no other proteins bind nonspecifically. The filter is overlayed upon the agar, and coated phage are allowed to bind to the antigen by way of their adherent antibody molecules. Filters are washed to remove nonspecifically bound phage.

30 Specifically bound phage therefore represent phage

-31-

encoding antibodies with the desired specificity. These can now be propagated by reinfection of bacteria.

Thus the present invention makes it possible to produce antigen-binding molecules which, like antibodies produced by presently-available techniques, bind to a selected antigen (i.e., having binding specificity). Antibodies produced as described can be used, for example, to detect and neutralize antigens and deliver molecules to antigenic sites.

10    EXAMPLE I    Amplification of IgM Heavy Chain Variable Region DNA from mRNA

IgM heavy chain variable DNA is amplified from mRNA by the procedure represented schematically in Figure 2. In Figure 2, Panel A depicts the relevant regions of the 15 poly adenylated mRNA encoding the secreted form of the IgM heavy chain. In Panel A, S denotes the sequences encoding the signal peptide which causes the nascent peptide to cross the plasma membrane, a necessary step in the processing and secretion of the antibody. V, D and J derive from separate exons and together comprise the 20 variable region.  $C_H^1$ ,  $C_H^2$ , and  $C_H^3$  are the three constant domains of  $C_H$ . "Hinge" encodes the hinge region. C, B and Z are oligonucleotide PCR primers used in the amplification process. The only constraints on Primers B 25 and Z are that they are complementary to the mRNA, and occur in the order shown relative to C. Primer C, in addition to being complementary to mRNA, has an extra bit of sequence at its 5' end which allows the cloning of its PCR product. This is described below. Panel B depicts 30 the reverse transcript DNA product of the mRNA primed by

- 32 -

oligonucleotide Z, with the addition of poly-dC by terminal transferase at the 3' end of the product. Panel C depicts the annealing of primer A to the reverse transcript DNA represented in Panel B. Primer A contains the restriction endonuclease site RE1, with additional DNA at its 5' end. The constraints on the RE1 site are described in Example 2. Panel D depicts the final double stranded DNA PCR product made utilizing primers A and B. Panel E depicts the PCR product shown in Panel D annealed to Primer C. Panel F is a blow up of panel E showing the structure of primer C. Primer C consists of two parts: a 3' part complementary to IgM heavy chain mRNA as shown, and a 5' part which contains restriction site RE2 and spacer. Constraints on RE2 are described in Example 2. Panel G depicts the final double stranded DNA PCR product utilizing Primers A and C and the product of the previous PCR (depicted in Panel D) as template. The S, V, D, J regions are again depicted.

EXAMPLE 2 Construction of Heavy Chain Framework Vector  
20 pFHC

A heavy chain framework vector, designated pFHC, is constructed, using known techniques (See Figure 3). It is useful for introducing antibody-encoding DNA into host cells, in which the DNA is expressed, resulting in antibody production. The circular plasmid (above) is depicted linearized (below) and its relevant components are shown. The neomycin antibiotic resistance gene ( $neo^R$ ) is useful for selecting transformed animal cells (Sambrook, J., et al., Molecular Cloning: A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold

-33-

Spring Harbor, NY (1989)). The bacterial replication origin and ampicillin antibiotic resistance genes, useful respectively, for replication in E. coli and rendering E. coli resistant to ampicillin, can derive from any number 05 of bacterial plasmids, including PBR322 (Sambrook, J., et al., Molecular Cloning: A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989)). The  $C\mu$  enhancer, which derives from the 10 intron between exons J and  $C_H^1$  of the  $C\mu$  gene, derives from any one of the cloned  $C\mu$  genes (Kawakami, T., et al., Nucleic Acids Research, 8:3933 (1980); Honjo, T., Ann. Rev. Immunol., 1:499 (1983)) and increases levels of transcription from antibody genes. LTR contains the 15 viral promoter from the Moloney MLV retrovirus DNA (Mulligan, R.C., Experimental Manipulation of Gene Expression, New York Academic Press, p. 155 (1983)). D represents the PCR primer described in the text, depicted in its 5' to 3' orientation. The only constraints on D are its orientation, its complementarity to 20 pFHC and its order relative to the RE1 and RE2 cloning sites. Preferably, D is within 100 nucleotides of RE1. The cDNA cloning site contains restriction endonuclease sites RE1 and RE2, separated by spacer DNA which allows their efficient cleavage. The constraints on RE1 and RE2 25 are described below. The  $C\mu$  exons, as described in the text and literature, direct the synthesis of IgM heavy chain. Only part of  $C_H^1$  is present, as described below.  $C_H^3$  is chosen to contain the  $C\mu s$  region which specifies a secreted form of the heavy chain ((Kawakami, T., et al., 30 Nucleic Acids Research, 8:3933 (1980); Honjo, T., Ann. Rev. Immunol., 1:499 (1983)). Finally, pFHC contains

- 34 -

poly A addition and termination sequences which can be derived from the C $\mu$  gene itself (Honjo, T., Ann. Rev. Immunol., 1:499 (1983); Kawakami, T., et al., Nucleic Acids Research, 8:3933 (1980)). One potential advantage 05 of using the entire C $\mu$  gene is that in some host cell systems, a membrane bound and secreted form of IgM may be expressed (Granowicz, E.S., et al., J. Immunol., 125:976 (1980)).

The plasmid can be produced by combining the 10 individual components, or nucleic acid segments, depicted in Figure 3, using PCR cassette assembly (See below). Because the entire nucleotide sequence of each component is defined, the entire nucleotide sequence of the plasma is defined.

15 The constraints on RE1 are simple. It should be the sole cleavage site on the plasmid for its restriction endonuclease. The choice of RE1 can be made by computer based sequence analysis (Intelligenetics Suite, Release 5:35, Intelligenetics).

20 The constraints on RE2 are more complex. First, it must be the sole cleavage site on the plasmid for its restriction endonuclease, as described for RE1.

Moreover, the RE2 site must be such that when the PCR 25 product is inserted, a gene is thereby created which is capable of directing the synthesis of a complete IgM heavy chain. This limits the choices for RE2, but the choices available can be determined by computer based sequence analysis. The choices can be determined as follows. First, a list of restriction endonucleases that 30 do not cleave pFHC is compiled (see Table 1).

- 35 -

TABLE 1

Non-Cutting Enzymes for the Mouse C<sub>μ</sub> Gene

	AatII	AhaII	AseI
	AvrII	BgII	BspHI
05	BssHII	BstBI	ClaI
	DraI	EagI	EcoRI
	EcoRV	FspI	HgaI
	HincII	HpaI	KpnI
	MluI	NaeI	NarI
10	NdeI	NotI	NruI
	PaeR7I	PvuI	RsrII
	SacII	Sall	Scal
	SfII	SnabI	SpeI
	SphI	SspI	StuI
15	Tth11I	XbaI	XbaI

These are called the "rare non-cutters." Next, the sequence of C<sub>H</sub>1 is rewritten with "N" at the third position of each codon and entered into the computer. This is called the "N-doped sequence" (See Figure 4). 20 Next, the rare non-cutters are surveyed by computer analysis for those which will cleave the N-doped sequence. The search program will show a possible restriction endonuclease site, assuming a match between N and the restriction endonuclease cutting site. For 25 example, with 39 rare non-cutters, 22 will cleave the N-doped sequence of C<sub>μ</sub> C<sub>H</sub>1, many of them several times (see Table 2). In this table, "Def" means a definite cut

- 36 -

site, of which there are none, because of the Ns. "Pos" means a possible cleavage site at the indicated nucleotide position if N is chosen appropriately. "Y" indicates any pyrimidine, "R" indicates any purine and 05 "N" indicates any nucleotide. The nucleotide positions refer to coordinates represented in Figure 4.

- 37 -

TABLE 2

	ENZYME	RECOGNITION	CUT SITE		
05	AatII	(GACGTC)	Def : none		
	AhaII	(GRCGYC)	Pos : 250	309	
10	AvrII	(CCTAGG)	Def : none		
	BspHI	(TCATGA)	Pos : 247	306	
15	BssHII	(GCGCGC)	Def : none		
	EcoRI	(GAATTC)	Pos : 204		
20	EcoRV	(GATATC)	Def : none		
	HgaI	(GACGCNNNNN) (NNNNNNNNNNNGCGTC)	Pos : 138		
25	HincII	(GTYRAC)	Def : none		
	HpaI	(GTTAAC)	Pos : 189		
30	KpnI	(GGTACC)	Def : none		
	NruI	(TCGCGA)	Pos : 284		
35	PaeR7	(CTCGAG)	Def : none		
	PvuI	(CGATCG)	Pos : 195	334	
40	ScaI	(AGTACT)	Def : none		
	SpeI	(ACTAGT)	Pos : 190	214	
45	SphI	(GCATGC)	Def : none		
	SspI	(AATATT)	Pos : 174	220	
	StuI	(AGGCCT)	Def : none		
	Tth111I	(GACNNNGTC)	Pos : 193	303	
	XbaI	(TCTAGA)	Def : none		
	XbaI	(CTCGAG)	Pos : 338		
			Def : none		
			Pos : 167	284	
			Pos : 359		
			Pos : 338		
			Pos : 371		
			Def : none		
			Pos : 149		
			Def : none		
			Pos : 212		
			Def : none		
			Pos : 338		
			Def : none		
			Pos : 190	339	

- 38 -

Most of these cleavage sites (about 60%) are compatible with the amino acids specified by  $C_H 1$ . Therefore, it is possible to mutate  $C_H 1$  to create a unique site for such an enzyme without altering the amino acid sequence

05 incoded by  $C_H 1$ . One sequence which illustrates this is shown below:

1)	...ala	met	gly	cys	leu	ala	arg	asp...
2)	...GCC	ATG	GGC	TGC	CTA	GCC	CGG	GAC...
3)	...GCC	ATG	GGC	TGC	CTA	<u>GCG</u>	<u>CGC</u>	GAC...

10

-----

BssHII

Line 1 represents part of the actual amino acid sequence specified by the mouse  $C_H 1$  gene region, and line 2 is the actual nucleotide sequence. By changing the sequence to the indicated nucleotides underlined on 15 line 3, a cleavage site for the rare non-cutter BssHII is created. The new sequence (containing the BssHII site) GCG CGC still encodes the identical amino acid sequence. Therefore, the sequence of the primer C is chosen to be the complement of line 3, and RE2 is the BssHII site.

20 Such a primer will function in the PCR and vector construction as desired. Other examples are possible, and the same process can be used in designing vectors and primers for cloning light chain variable regions.

The choice for primer C puts a constraint on pFHC. 25 In the example shown, the  $C_H 1$  region contained on pFHC must begin at its 5' end with the mutant sequence GCG CGC. Such mutant fragments can be readily made by the process of PCR cassette assembly described below.

- 39 -

The process of PCR cassette assembly is a method of constructing plasmid molecules (in this case the plasmid pFHC) from fragments of DNA of known nucleotide sequence. One first compiles a list of restriction endonucleases

05 that do not cleave any of the fragments. Each fragment is then individually PCR amplified using synthesized oligonucleotide primers complementary to the terminal sequences of the fragment. These primers are synthesized to contain on their 5' ends restriction endonuclease

10 cleavage sites from the compiled list. Thus, each PCR product can be so designed that each fragment can be assembled one by one into a larger plasmid structure by cleavage and ligation and transformation into E. coli. Using this method, it is also possible to make minor

15 modifications to modify the terminal sequence of the fragment being amplified. This is done by altering the PCR primer slightly so that a mismatch occurs. In this way it is possible to amplify the C $\mu$  gene starting precisely from the desired point in C<sub>H</sub>1 (as determined by

20 oligo C above) and creating the RE2 endonuclease cleavage site.

-40-

CLAIMS

1. An in vitro process for synthesizing DNA encoding a family of antigen-combining proteins, comprising the steps of:
  - 05 a) obtaining DNA containing genes encoding antigen-combining proteins;
  - b) combining the DNA containing genes encoding antigen-combining proteins with sequence specific primers which are oligonucleotides homologous to conserved regions of the genes; and
  - 10 c) performing sequence specific gene amplification.
2. DNA encoding a family of antigen-combining proteins produced by the process of Claim 1.
- 15 3. The process of Claim 1 wherein sequence specific gene amplification is performed by the polymerase chain reaction.
4. The process of Claim 3 wherein the sequence specific primers are bidirectional.
- 20 5. The process of Claim 3 wherein the sequence specific primers are nested unidirectional primers.
6. The process of Claim 1 wherein the antigen-combining proteins are immunoglobulins.

-41-

7. The process of Claim 6 wherein the immunoglobulins are selected from the group consisting of heavy chains and light chains.
8. The process of Claim 7 wherein the heavy chains are  $\mu$  chains.  
05
9. The process of Claim 1 wherein the DNA containing genes encoding antigen-combining proteins is cDNA of RNA from antibody-producing cells.
10. The process of Claim 1 wherein the DNA containing genes encoding antigen-combining proteins is genomic DNA from antibody-producing cells.  
10
11. The process of Claim 8 wherein the antigen-combining proteins are of mammalian origin.
12. The process of Claim 1 wherein the primers are oligonucleotides homologous to conserved regions of the constant regions of immunoglobulin genes.  
15
13. The process of Claim 1 wherein the primers are oligonucleotides homologous to the conserved regions of the variable regions of immunoglobulin genes.
- 20 14. The process of Claim 1 wherein the primers contain at least one restriction endonuclease cloning site.
15. The process of Claim 1 wherein the primers are selected from the group consisting of oligonucleotide B of Figure 2 and oligonucleotide C of Figure 2.  
25

-42-

16. A method of creating a diverse starter library of DNAs encoding families of antigen-combining proteins comprising cloning the product of Claim 1 into an appropriate vector.
- 05 17. A diverse starter library of DNAs encoding families of antigen-combining proteins produced by the method of Claim 14.
18. The method of Claim 16 wherein the vector is a prokaryotic vector or a eukaryotic vector.
- 10 19. The method of Claim 16 wherein the vector is a viral vector or a retroviral vector.
20. The method of Claim 16 wherein the vector is a plasmid.
21. The method of Claim 20 wherein the plasmid is selected from the group consisting of pFHC and pLHC.
- 15 22. The method of Claim 16 wherein the vector is selected from the group consisting of expression vectors and cloning vectors.
- 20 23. The method of Claim 22 wherein the expression vector is appropriate for expression of the variable region of an antigen-combining protein as a chimeric molecule in register with a framework protein.
24. The method of Claim 23 wherein the framework protein is an immunoglobulin.

-43-

25. The method of Claim 24 wherein the immunoglobulin is all or a portion of the constant region of the  $\mu$  heavy chain.
- 05 26. The method of Claim 16 further comprising creating a collection of viral particles from viral vector-based libraries of DNA encoding antigen-combining proteins by the process of introducing viral vectors into host cells in which they replicate and form viral particles.
- 10 27. A method of producing a high diversity library of DNA encoding families of antigen-combining proteins comprising mutagenizing the product of Claim 16.
- 15 28. A high diversity library of DNA encoding families of antigen-combining proteins produced by the method of Claim 27.
29. The method of Claim 27 wherein mutagenizing is carried out by random chemical mutagenesis.
- 20 30. The method of Claim 27 wherein mutagenizing is carried out by performing the polymerase chain reaction under limiting nucleotide conditions.
31. The method of Claim 27 wherein mutagenizing is carried out in such a manner that mutagenesis is limited to DNA encoding variable regions of the antigen-combining protein.
- 25 32. A process of producing a diverse population of host cells which comprises introducing into host cells

-44-

DNA of the starter library or high diversity  
libraries of antigen-combining proteins.

33. Host cells produced by the method of Claim 32.
- 05 34. The process of Claim 32 wherein the host cells are prokaryotic.
35. The process of Claim 32 wherein the host cells are eukaryotic.
- 10 36. The process of Claim 35 wherein the host cells are selected from the group consisting of immortalized cultured mammalian cells.
37. The process of Claim 36 wherein the immortalized cultured mammalian cells are selected from the group consisting of myelomas and plasmacytomas.
- 15 38. The process of Claim 32 wherein the libraries encoding families of antigen-combining proteins are introduced into host cells by a method selected from the group consisting of: electroporation, calcium phosphate coprecipitation, protoplast fusion, viral infection, and cell fusion.
- 20 39. The process of Claim 32 wherein the libraries of DNAs encoding families of antigen-combining proteins is contained in an expression vector.
- 25 40. The process of Claim 32 wherein the DNAs encoding families of antigen-combining proteins encode antigen-combining proteins selected from the group

-45-

consisting of immunoglobulin heavy chain variable regions or immunoglobulin light chain variable regions.

41. The process of Claim 40 wherein DNAs encoding immunoglobulin heavy chain variable regions are introduced simultaneously with or sequentially to DNAs encoding immunoglobulin light chain variable regions.  
05
42. The method of Claim 32 further comprising identifying cells which produce antigen-combining molecules of selected specificity.  
10
43. The method of Claim 42 wherein identifying of cells which produce antigen-combining molecules of selected specificity is carried out by assaying cellular supernatants for antigen-combining activity.  
15
44. The method of Claim 42 wherein identifying of cells which produce antigen-combining molecules of selected specificity is carried out by a nitrocellulose filter overlay technique.  
20
45. The method of Claim 44 wherein cells producing antigen-combining molecules of selected specificity are enriched for cells producing antigen-combining molecules on their surface by affinity matrix chromatography.  
25
46. Cells produced by the method of Claim 42.

- 46 -

47. Antigen-combining molecules produced by cells of  
Claim 42.
48. DNAs encoding immunoglobulin heavy chain variable  
regions or immunoglobulin light chain variable  
05 regions, present in cells of Claim 42.
49. Viruses produced by the method of Claim 26.
50. A method of isolating viruses of Claim 49 encoding  
antigen-combining molecules of selected specificity,  
comprising the steps of:  
10 a) infecting host cells with an appropriate virus  
containing DNA encoding antigen-combining molecules;  
b) coating the virus with antigen-combining  
molecules which the virus encodes; and  
c) subjecting the product of step (b) to  
15 affinity-matrix selection, to separate the virus  
according to the antigen-combining molecules they  
contain.\*
51. Viruses produced by the method of Claim 50.
52. Antigen-combining molecules encoded by viruses of  
20 Claim 51.

1/5

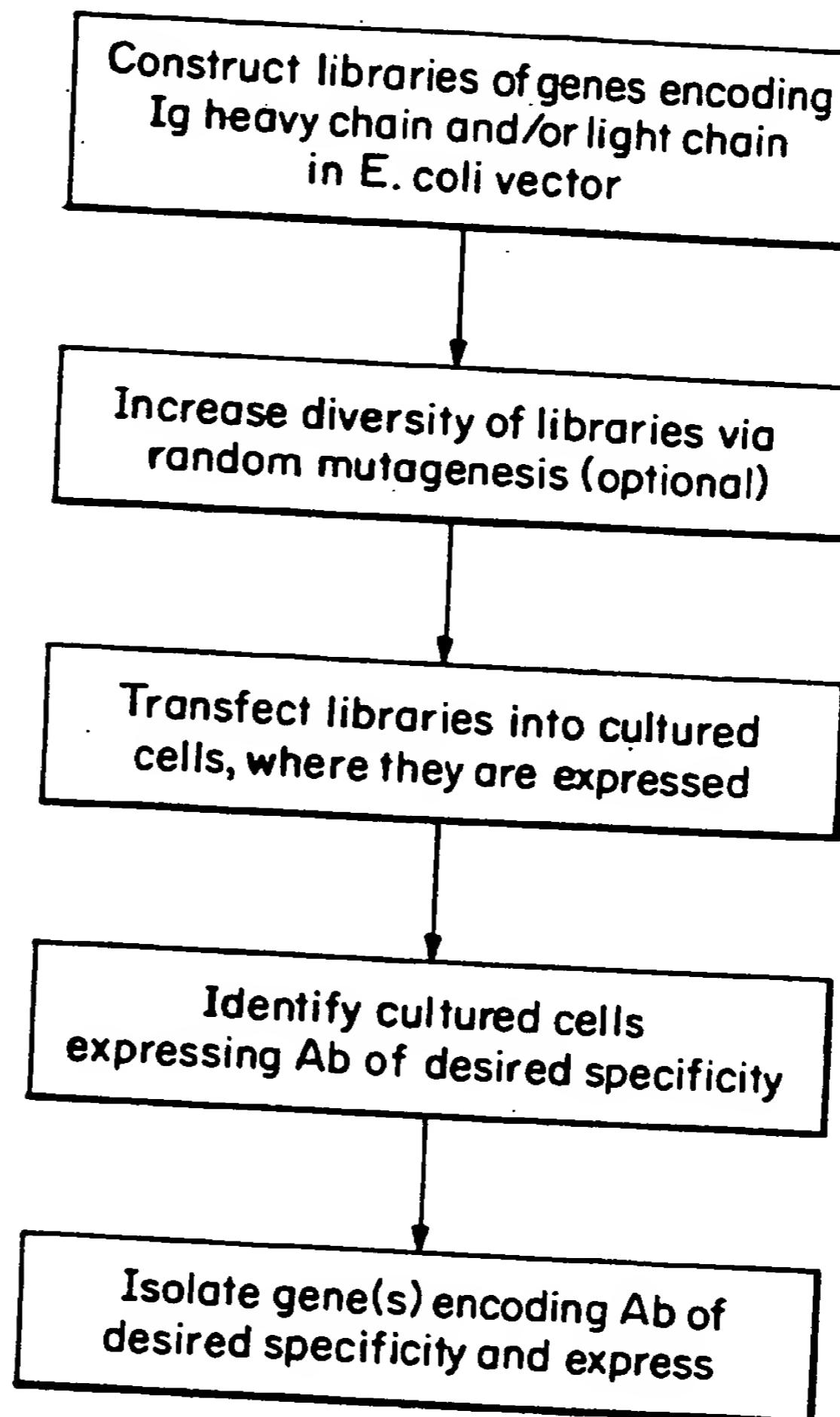


Fig. 1

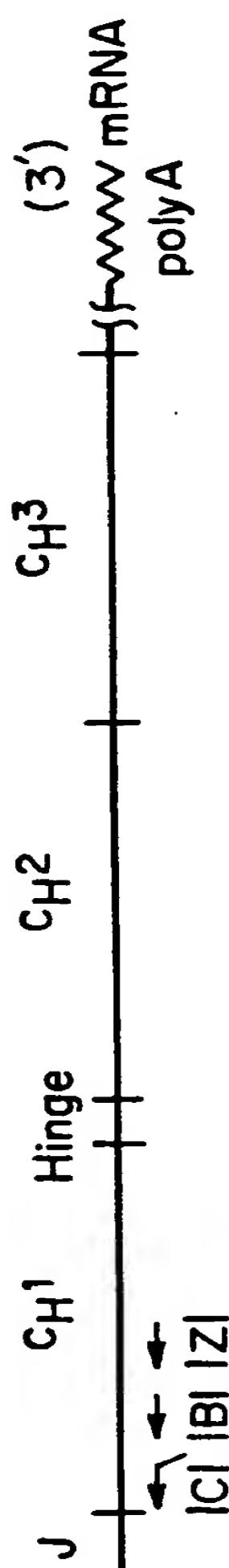


Fig. 2A (5') S V D J CH<sub>1</sub> Hinge CH<sub>2</sub> CH<sub>3</sub> (3') CCC...CCCC polyA

Fig. 2B (3') CCC...CCCC

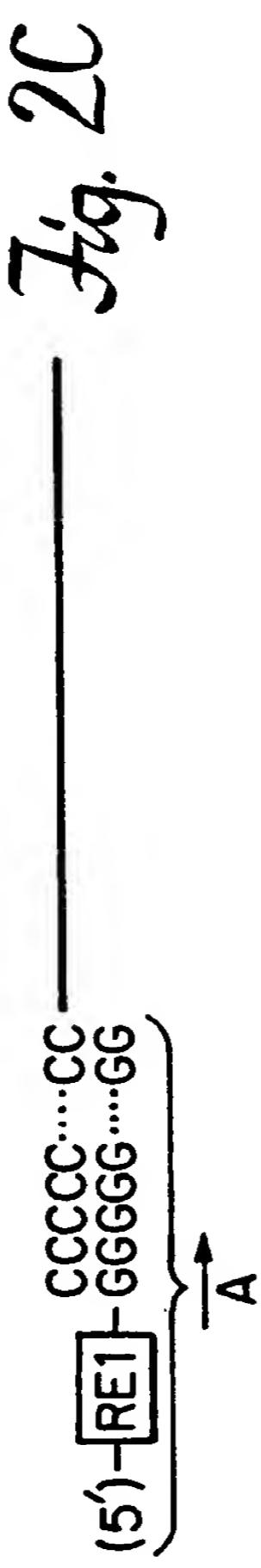


Fig. 2C (5')-RE1-CCCC...CC

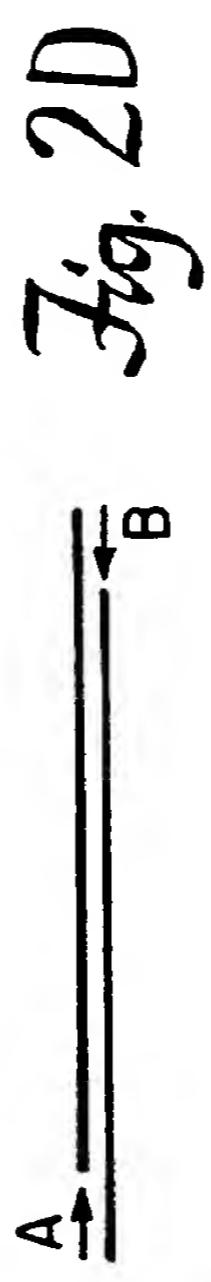


Fig. 2D A B

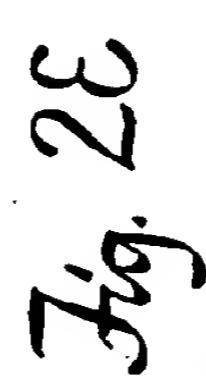


Fig. 2E

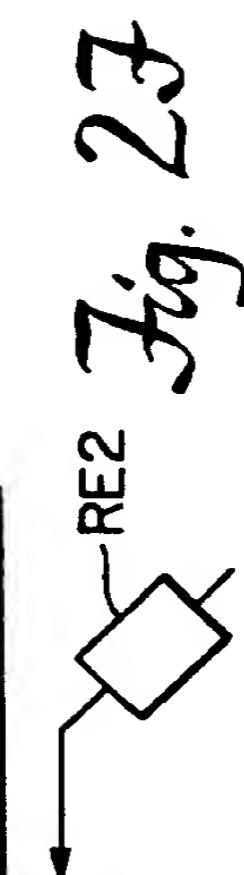


Fig. 2F

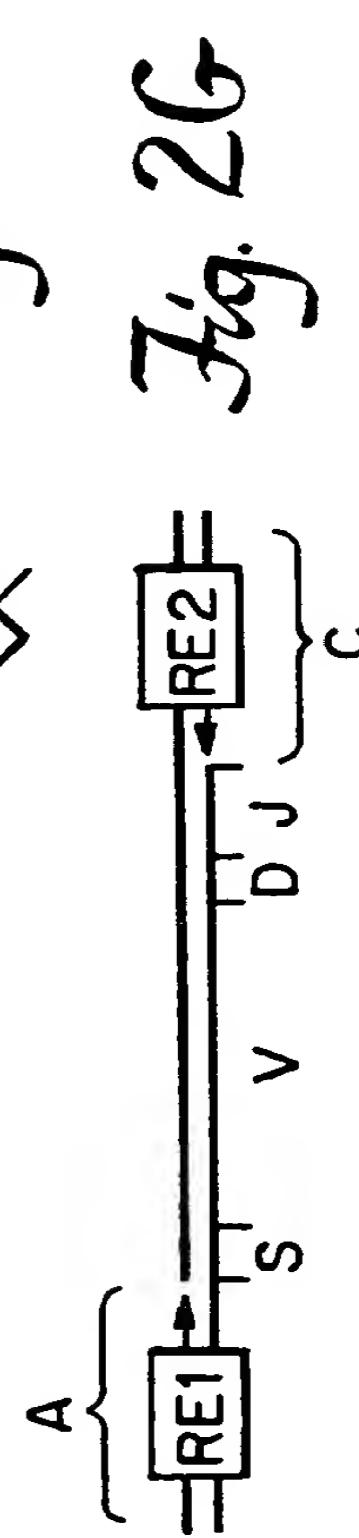


Fig. 2G

3/5

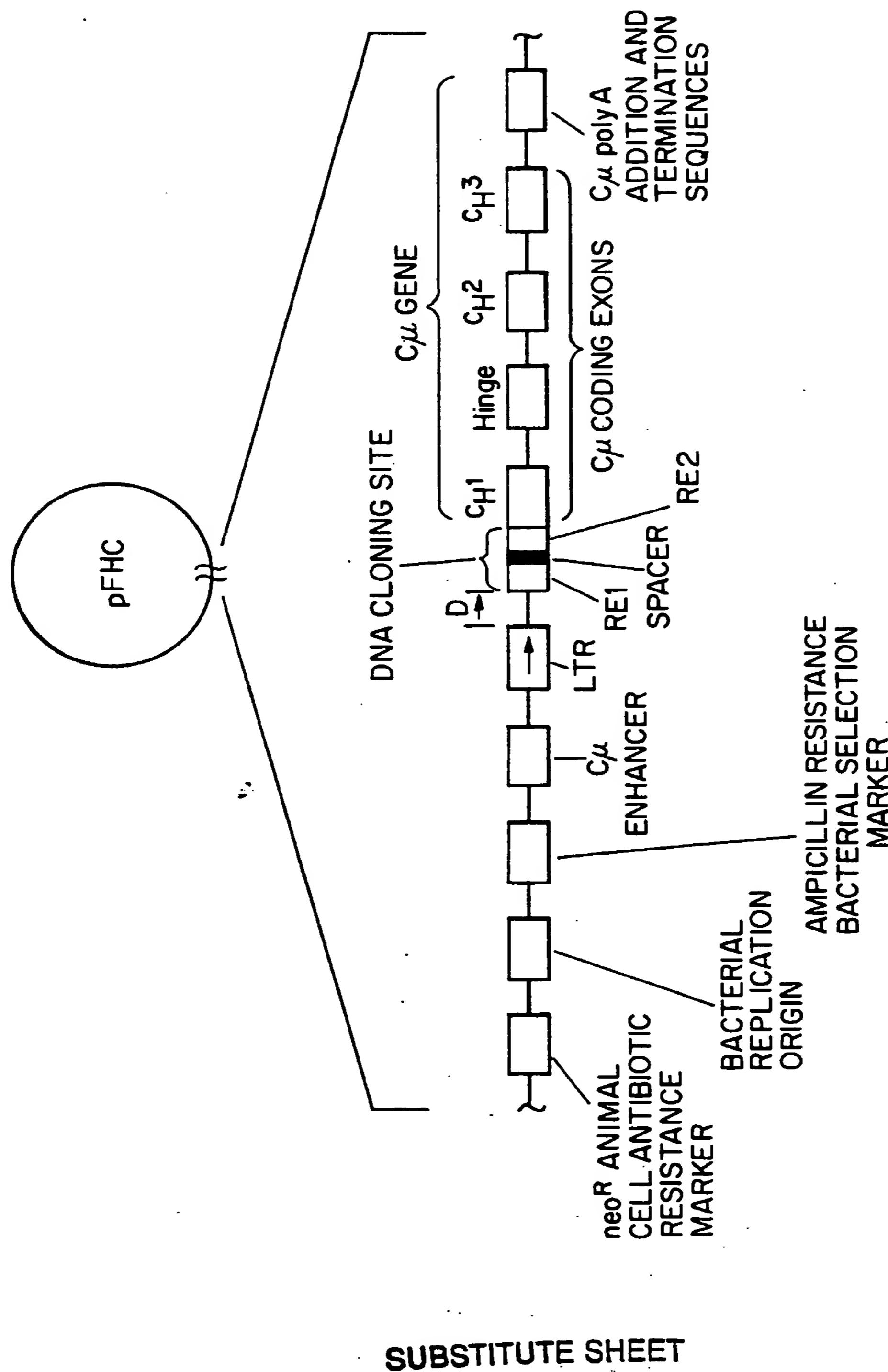


Fig. 3

## **SUBSTITUTE SHEET**

## **SUBSTITUTE SHEET**

# INTERNATIONAL SEARCH REPORT

International Application No PCT/US 91/00209

## I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) <sup>4</sup>

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC<sup>5</sup>: C 12 N 15/13, C 07 K 15/28

## II. FIELDS SEARCHED

Minimum Documentation Searched <sup>7</sup>

Classification System	Classification Symbols
IPC <sup>5</sup>	C 07 K, C 12 N

Documentation Searched other than Minimum Documentation  
to the Extent that such Documents are Included in the Fields Searched <sup>8</sup>

## III. DOCUMENTS CONSIDERED TO BE RELEVANT<sup>9</sup>

Category <sup>10</sup>	Citation of Document, <sup>11</sup> with indication, where appropriate, of the relevant passages <sup>12</sup>	Relevant to Claim No. <sup>13</sup>
X	Science, vol. 246, 8 December 1989, W.D. Huse et al.: "Generation of a large combinatorial library of the immunoglobulin repertoire in phage lambda", pages 1275-1281 see the whole article; especially tables 1 and 2 --	1-3,5-14, 16-20,21- 25
X	Proceedings of the National Academy of Sciences, vol. 86, August 1989, L. Sastry et al.: "Cloning of the immunological repertoire in Escherichia coli for generation of monoclonal catalytic antibodies: construction of a heavy chain variable region-specific cDNA library", pages 5728-5732 see the whole article --	1-3,5-14, 16-20,21- 25
	..	

- Special categories of cited documents: <sup>10</sup>
- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "A" document member of the same patent family

## IV. CERTIFICATION

Date of the Actual Completion of the International Search

16th May 1991

Date of Mailing of this International Search Report

25/06/91

International Searching Authority

EUROPEAN PATENT OFFICE

Signature of Authorized Officer

Danielle van der Haas

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, <sup>11</sup> with indication, where appropriate, of the relevant passages	Relevant to Claim No.
X	Nature, vol. 341, no. 6242, 12 October 1989, (London, GB), E.S. Ward et al.: "Binding activities of a repertoire of single immunoglobulin variable domains secreted from Escherichia coli", pages 544-546 see the whole article --	1-3,5-14, 16-20,21-25
X	Proceedings of the National Academy of the USA, vol. 86, no. 10, May 1989, (Washington, DC, US), R. Orlandi et al.: "Cloning immuno-globulin variable domains for expression by the polymerase chain reaction", pages 3833-3837 see the whole article --	1-3,5-14, 16-20,21-25
A	Clinical Chemistry, vol. 35, no. 9, September 1989, G.P. Moore: "Genetically engineered antibodies", pages 1849-1853 see page 1852, left-hand column	1-52
-----		